Biochemical and structural characterization of CYP109A2, a vitamin D3 25-hydroxylase from Bacillus megaterium

Abdulmughni, Ammar; Jóźwik, Ilona K.; Brill, Elisa; Hannemann, Frank; Thunnissen, Andy-Mark W H; Bernhardt, Rita

Published in:
Febs Journal

DOI:
10.1111/febs.14276

Publication date:
2017

Biochemical and structural characterization of CYP109A2, a vitamin D_3 25-hydroxylase from *Bacillus megaterium*

MISS ILONA JÓŹWIK (Orcid ID : 0000-0002-7150-5802)

Ammar Abdulmughni[1#], Ilona K. Jóźwik[2#], Elisa Brill[1], Frank Hannemann[1], Andy-Mark W.H. Thunnissen[2Ψ*] and Rita Bernhardt[1*]

Author information:

[1] - Department of Biochemistry, Campus B2.2, 66123, Saarland University, Saarbrücken, Germany

[2] - Laboratory of Biophysical Chemistry, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 7, 9747 AG, Groningen, The Netherlands

[Ψ] – Current address: Molecular Enzymology group, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 4, 9747 AG, Groningen, The Netherlands

[#] - These two authors contributed equally to this study.

*** - Correspondence:**

A.-M. W. H. Thunnissen,
Molecular Enzymology group, Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 4, 9747 AG, Groningen, The Netherlands. Tel: +31 50 3634209, Fax: +31 50 3634165, E-mail: a.m.w.h.thunnissen@rug.nl

R. Bernhardt, Institute of Biochemistry, Saarland University, Campus B 2.2, 66123 Saarbrücken, Germany. Tel: +49 681 302 4241, Fax: +49 681 302 4739, E-mail: ritabern@mx.uni-saarland.de

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/febs.14276

Running title: Vitamin D_3 25-hydroxylase CYP109A2.

Article type : Original Article

Keywords: cytochrome P450, biocatalysis, vitamin D_3, *Bacillus megaterium*, regio-selectivity

Database: The atomic coordinates and structure factors have been deposited in the Protein Data Bank with accession code **5OFQ** (substrate-free CYP109A2).

Enzymes: Cytochrome P450 monooxygenase CYP109E1, **EC 1.14.14.1**, UniProt ID: **D5DF88**, Ferredoxin, UniProt ID: **D5DFQ0**, cytochrome P450 reductase, **EC 1.8.1.2**, UniProt ID: **D5DGX1**.

Abbreviations: VD_3: Vitamin D_3, 25(OH)VD_3: 25-hydroxyvitamin D_3, P450: cytochrome P450

Abstract

Cytochrome P450 enzymes are increasingly investigated due to their potential application as biocatalysts with high regio- and/or stereo-selectivity and under mild conditions. Vitamin D_3 metabolites are of pharmaceutical importance and are applied for the treatment of vitamin D_3 deficiency and other disorders. However, the chemical synthesis of vitamin D_3 derivatives shows low specificity and low yields. In this study, cytochrome P450 CYP109A2 from *Bacillus megaterium* DSM319 was expressed, purified and shown to oxidize vitamin D_3 with high regio-selectivity. The *in vitro* conversion, using cytochrome P450 reductase (BmCPR) and ferredoxin (Fdx2) from the same strain, showed typical Michaelis-Menten reaction kinetics. A whole-cell system in *B. megaterium* overexpressing CYP109A2 reached 76 ± 5% conversion after 24h and allowed to identify the main product by NMR analysis as 25-hydroxylated vitamin D_3. Product yield amounted to 54.9 mg L^{-1} day^{-1}, rendering the established whole-cell system as a highly promising biocatalytic route for the production of this valuable metabolite. The crystal structure of substrate-free CYP109A2 was determined at 2.7 Å resolution, displaying an open conformation. Structural analysis predicts that

CYP109A2 uses a highly similar set of residues for vitamin D_3 binding as the related vitamin D_3 hydroxylases CYP109E1 from *Bacillus megaterium* and CYP107BR1 (Vdh) from *Pseudonocardia autotrophica*. However, the folds and sequences of the BC loops in these three P450s are highly divergent, leading to differences in the shape and apolar/polar surface distribution of their active site pockets, which may account for the observed differences in substrate specificity and the regio-selectivity of vitamin D_3 hydroxylation.

Introduction

Vitamin D_3 (VD_3) is a secosteroid that is produced in the human skin from the precursor compound 7-dehydrocholesterol upon exposure to UV light irradiation, resulting in the opening of the precursor B-ring. Afterwards, VD_3 is metabolized successively in the liver and kidneys to form two derivatives that are highly important for human health, the 25-hydroxyvitamin D_3 ($25(OH)VD_3$) and 1α,25-dihydroxyvitamin D_3 ($1\alpha,25(OH)_2VD_3$, **Figure 1**). While the latter derivative is the biologically active form of VD_3, the former one is the most abundant circulating metabolite of VD_3 in the blood and is therefore considered to be an optimal indicator of the nutritional status concerning VD_3 in patients [1,2]. Maintaining an optimal level of $25(OH)VD_3$ is crucial for reducing the risk of several chronic diseases or cancer conditions [3]. Moreover, the $25(OH)VD_3$ metabolite is acknowledged for better absorption properties as compared to VD_3 in certain supplement-based treatments [4,5]. Unfortunately, chemical synthesis of VD_3 derivatives is a challenging, multistep procedure resulting in low product yields, limiting their usefulness as therapeutic dietary supplements [6]. Biotechnological approaches involving enzymes capable of specifically oxidizing VD_3 offer an attractive alternative to circumvent such limitations.

In humans, the two consecutive VD_3 hydroxylation steps, at carbons C-25 and C-1, are catalyzed by proteins belonging to the cytochrome P450 superfamily (P450s or CYPs) [7]. The P450s are heme-containing enzymes that catalyze the monooxygenation of a large variety of organic substrates like steroids, fatty acids, and are generally known to be involved in the oxidative

metabolism of xenobiotics and drugs. Considering the broad range of P450 substrates, the enzymes are of benefit to synthesis of pharmaceutically and biotechnologically valuable compounds, in particular when high regio- and stereo-selectivity of hydroxylation is required. Microbial P450s are favored, as they are soluble and more stable than their eukaryotic counterparts [8-10]. Thus far, only a few bacterial P450 enzymes have been reported that hydroxylate VD_3 to produce the desired 25(OH)VD_3 metabolite, like CYP105A1 from *Streptomyces griseolus* [11] or CYP107BR1 (Vdh) from *Pseudonocardia autotrophica* [12]. Most recently, our investigations led to the identification and characterization of another bacterial VD_3-hydroxylase, cytochrome P450 CYP109E1 from *Bacillus megaterium* DSM319 [13]. The improvement of activity and/or regio-selectivity of the three above-mentioned enzymes by protein engineering was supported by the knowledge of their structure-function relationships, thanks to determination of their crystal structures in both the open, substrate-free and closed, substrate-bound states [14-18].

Here we describe the identification and analysis of a new cytochrome P450 enzyme from *Bacillus megaterium* DSM319, classified as CYP109A2, which shares 45% amino acid sequence identity with CYP109E1. Similar to CYP109E1, CYP109A2 exhibits 25-hydroxylation activity towards VD_3, however, with significantly higher regio-selectivity. A whole-cell *B. megaterium*-based VD_3 oxidation system was established for CYP109A2 and resulted in higher yields of 25(OH)VD_3 as compared to the previously reported most productive variant of CYP109E1, *i.e.*, mutant I85A [13]. The three-dimensional structure of CYP109A2 was obtained by X-ray crystallography and compared to the structures of other bacterial P450s that convert VD_3, suggesting putative structural features associated with differences in substrate specificity and regio-selectivity.

Results and Discussion

Identification, sequence analysis and initial characterization of CYP109A2

Bioinformatic mining of the *Bacillus megaterium* DSM319 genome previously led to identification of four genes putatively encoding for cytochrome P450 enzymes (CYP102A1, CYP106A1, CYP109E1 and CYP109A2) [18,19]. Here we focus on CYP109A2, the remaining P450 that has not been studied

in detail yet. Its open reading frame (KEGG: BMD_2035) encodes for a 403 amino acid long protein containing three characteristic sequence motifs, which support its annotation as a P450 enzyme: (i) heme-binding domain signature proceeding the L helix (FxxGx(H/R)xCxG), (ii) the acid-alcohol pair in the I helix (Glu244, Thr245) and (iii) ExxR motif in the K helix (**Figure 2**). Interestingly, CYP109A2 shares 45% sequence identity with CYP109E1 from the same organism (the sequence identity with CYP106A1 and CYP102A1 is 36% and 24%, respectively). Comparative sequence analysis including proteins from different bacterial sources further shows that CYP109A2 shows significant sequence homology to VD_3 hydroxylating P450s from the CYP107 and CYP105 family (37% identity to CYP107BR1 (Vdh) from *P. autotrophica*, 36% to CYP107CB2 from *B. lehensis* G1, 32% identity to CYP105A1 from *Streptomyces griseolus* and 30% to CYP105A2 from *P. autotrophica*). Residues in CYP109A2 with the highest degree of conservation have a role in heme binding or belong to substrate recognition site 4 (SRS4) (central part of I helix), whose residues have common roles in P450s associated with substrate binding, oxygen activation and proton delivery during catalysis [20]. Other regions associated with substrate binding (SRS2, SRS5 and SRS6) are less well conserved, although relatively higher conservation levels may be observed in SRS5 or SRS6 upon pairwise comparison of P450s belonging to the same CYP family, like CYP109A2 and CYP109E1. The lowest degree of conservation is displayed by the amino acids belonging to SRS1 (the BC loop region) and SRS3 (the G helix) (**Figure 2**).

Since a significant sequence similarity between bacterial P450s belonging to the same CYP family may correlate to similarities in substrate scope [21], we set out to establish whether VD_3, being a substrate for CYP109E1, is also converted by CYP109A2. To produce sufficient amounts of pure and functional CYP109A2 enzyme for the intended functional analysis, the *cyp109a2* gene was cloned into the pET17b vector and expressed in *E. coli* C43 (DE3) cells with high yield (4210 $nmol*L^{-1}$). The protein was purified by affinity chromatography and displayed characteristic P450 spectroscopic behavior, considering both the UV-visible spectrum of its oxidized ligand-free and reduced CO-bound forms (Figure 3). Spectroscopic analysis thus confirmed that the purified protein contained the correctly incorporated heme group and may be functional as a P450 enzyme.

CYP109A2 catalyzes regio-selective oxidation of vitamin D_3 *in vitro*

Cytochromes P450 are external monooxygenases and thus require electrons for their catalytic activity from different (autologous or heterologous) redox partners [22]. Selection of suitable redox partners to support CYP109A2 activity was based on previous observations that *B. megaterium* cytochrome P450 reductase (BmCPR) and a ferredoxin (Fdx2) were able to support CYP106A1 with much higher efficiency than the commonly used AdR-$Adx_{4\text{-}108}$ (adrenodoxin reductase – adrenodoxin) redox partners [23]. Indeed, reconstitution of CYP109A2 activity was accomplished using both redox partner systems, but the CYP109A2-catalysed *in vitro* reactions (200 µM, 1h) yielded a significantly higher VD_3 conversion ratio when using the BmCPR-Fdx2 redox pair (36 ±4 %) (**Figure 4A**) as compared to the AdR-$Adx_{4\text{-}108}$ pair (20 ± 5 %, data not shown). Usage of the different redox pairs (BmCPR-Fdx2 or AdR-$Adx_{4\text{-}108}$) to reconstitute CYP109A2 activity towards VD_3 resulted in identical product patterns, showing one main product P1 (~ 64% of total products) and one minor product P2 (~ 36% of total products) (**Figure 4A**). Determination of the steady-state kinetics of VD_3 conversion by CYP109A2 showed a typical Michelis-Menten behavior with k_{cat} nd K_M values of 3.42 ± 0.15 min^{-1} and 49.0 ± 5.8 µM, respectively (**Figure 5**). These results demonstrate that the CYP109A2 *in vitro* activity towards VD_3 is lower than that of other previously characterized bacterial VD_3 hydroxylases [24]. In addition, CYP109A2 displays a lower substrate conversion ratio in comparison to its most closely related homolog CYP109E1, which was shown to convert ~90 % of VD_3 under the same conditions (when supported by the AdR- $Adx_{4\text{-}108}$ pair) [13]. The affinity for VD_3 is also lower in CYP109A2 as compared to wild type CYP105A1 or Vdh, which showed K_M values of 0.54 ± 0.09 µM and 13.5 ± 3.7 µM, respectively [12,25].

On the other hand, the regio-selectivity of CYP109A2, as evident from its product pattern of VD_3 conversion, is superior to that of the other bacterial VD_3 hydroxylases. CYP109A2 converts VD_3 into two products only, whereas previously it was shown that CYP109E1 converts the same substrate to seven products [13]. CYP105A1 and Vdh convert VD_3 to three products [12,25], while for the VD_3

25-hydroxylase CYP105A2 from *P. autotrophica* the exact product profile is not yet determined [24]. Interestingly, upon comparison of the CYP109A2 product retention times with the available authentic standards and HPLC data previously obtained for CYP109E1, it was hypothesized and later shown (see below) that product P1 is 25-hydroxyvitamin D_3. Therefore, to the best of our knowledge the newly identified enzyme, CYP109A2 from *B. megaterium*, is the most regio-selective bacterial VD_3 hydroxylase characterized so far.

Whole-cell based bioconversion of vitamin D_3 by CYP109A2

Our recent study of CYP109E1 has demonstrated that the *B. megaterium*-based whole cell system may be an efficient and sustainable route to large-scale production of VD_3 metabolites like 25(OH)VD_3 [13]. *Bacillus megaterium* is a Gram-positive, non-pathogenic bacterium and a well-studied biotechnologically applicable production host. Its use in the pharmaceutical and food industries is preferred thanks to the ability of this bacterium to grow on a variety of carbon sources and simple media as well as due to its high protein production capacity and plasmid stability [26,27]. Importantly, as shown by research results considering various P450 enzymes, the *B. megaterium* offers certain advantages, of being not only a source of new, uncharacterized P450s but also a very good host for P450s and/or their homologous or heterologous redox partners [19,28-30]. Use of the whole-cell expressed P450(s) in a biotransformation process allows for regeneration of the expensive NADPH cofactor that is necessary for P450 activity.

The *B. megaterium* MS941 cells were transformed with expression plasmid pSMF2.1.CYP109A2, and using such a system 76 ± 5% of 200 µM VD_3 was successfully converted within 24 h into one main product P1 (90%) and one minor product P2 (10%) (**Figure 4B**). In comparison to the profile observed *in vitro* the product ratio obtained *in vivo* is slightly changed, favoring generation of the main product P1 over product P2. The changes in the product profile are

most probably caused by the fact that no heterologous redox partners were expressed along with CYP109A2, leaving only the endogenous *B. megaterium* redox proteins (which remain to be identified) to support its activity. In fact, it was recently observed that varying the redox partners for reconstitution of P450-based reactions can influence the observed product ratios [31,32].

Product isolation and identification by nuclear magnetic resonance analysis

As for the *in vitro* conversion of VD_3 by CYP109A2, the main hydroxylated product obtained by the *in vivo* conversion showed an identical retention time in HPLC analysis as the authentic standard of 25-hydroxyvitamin D_3. To allow full characterization of the product(s) by nuclear magnetic resonance spectroscopy (NMR), the *in vivo* reactions were up-scaled to 400 ml of total culture volume (200 µM VD_3), and thus 16 mg of the product of interest (P1) were obtained. Unfortunately, the yields of product P2 were not sufficient to allow for the direct product identification by NMR. For the isolated product P1 the ^{1}H NMR and ^{13}C NMR spectra were recorded. It was observed that in comparison to standard VD_3, product P1 showed resonances for an additional tertiary hydroxyl group (at C-25) in the ^{13}C NMR spectrum (δ_C 71.15), whereas the resonances of the methyl groups C-26 and C-27 both appeared as singlets (δ_H 1.19 s, 6H) in the ^{1}H NMR spectrum. This clearly indicated that CYP109A2 catalyzes conversion of VD_3 to 25(OH)VD_3 as the main product (**Figure 6**). The NMR data are in accordance with reported literature (Table S1) [33,34].

Efficient biocatalytic production of 25-hydroxyvitamin D_3 by CYP109A2

Based on the results of the *in vivo* conversion of VD_3 performed by the CYP109A2-based whole-cell system, the yield of 25(OH)VD_3 was calculated to be 54.9 mg L^{-1} day^{-1}. Hence, it becomes apparent that the production yield of the CYP109A2-based whole-cell system is significantly higher as compared to other bacterial *in vivo* CYP-based systems known to produce the same product. For example, the previously studied CYP109E1 expressed in *B. megaterium* under the same conditions

and reaction time produced 24.5 mg L^{-1} day^{-1} of 25(OH)VD_3, whereas with its best mutant I85A, the achieved yield was 45 mg L^{-1} day^{-1} [13]. The yield achieved with the CYP109A2-based system is also higher than the yield reported for the most productive mutant of CYP105A1 (R73A/R84A), which produced 7.8 mg 25(OH)VD_3 L^{-1} day^{-1} [14]. Only Vdh expressed in *A. autotrophica* (now referred to as *P. autotrophica*) was reported to give a higher yield than CYP109A2 being 137 mg L^{-1} [35]. However, this value was obtained after 72 h, whereas, in case of CYP109A2 described here 54.9 mg L^{-1} were obtained already after 24 h. Thus, the *B. megaterium* CYP109A2-based whole-cell system offers a very promising biotechnological alternative to production of 25-hydroxyvitamin D_3 in comparison to other characterized P450 biocatalysts.

Crystal structure of substrate-free CYP109A2

In order to determine the geometry of its active site pocket and predict the residues responsible for VD_3 binding, the CYP109A2 enzyme was studied by X-ray crystallography. For crystallization, a C-terminally His-tagged protein was purified using the previously described three-step purification procedure, resulting in protein sample of >98% purity and low polydispersity (12.2 %) as confirmed by SDS-PAGE analysis and dynamic light scattering measurements (data not shown). The substrate-free CYP109A2 crystallized in an orthorhombic space group ($P2_12_12_1$), with four monomers in the asymmetric unit and a solvent content of 50%. The crystal structure was solved at 2.7 Å resolution by the molecular replacement method, using the previously determined structure of CYP109E1 (sequence similarity 45%, PDB entry 5L90, [18]) as a search model (see Table 1 for data collection and refinement statistics). All four CYP109A2 molecules have well defined electron density, except for the fifteen or sixteen N-terminal residues and the C-terminal His_6-tags, for which no electron density was observed. The central parts of the BC loop, FG loop and N-terminal end of the G helix show relatively high B factors, related to high mobility of those regions, as generally observed in P450 crystal structures. The four CYP109A2 molecules found in the asymmetric unit have almost identical conformations (the root mean square deviation [r.m.s.d] of Cα-backbone atoms ranges between 0.36-0.43 Å). Therefore, only one of the protein molecules will be described further. The

structure of CYP109A2 (residues 16-401) exhibits the typical organization of a P450 domain, containing fifteen helices (labeled A-L, A', B' and K'), and ten β-strands (arranged in three β-sheets; in the same structural arrangement as observed in the CYP109E1 crystal structure, **Figure 7A**). The centrally located heme group is sandwiched between the I and L helices, similarly as in other P450s. The bound heme is covalently attached to the protein by an iron-sulfur bond with Cys351, and is further stabilized by several residues providing van der Waals and hydrophobic interactions, while its propionate side chains interact with the side chains of His91, Arg95, Arg293 and His349. At the distal side of the heme, no clear electron density was observed for a heme-bound water molecule, in contrast to many other P450 crystal structures determined in the substrate-free state. A small blob of electron density is located at approximately 5 Å distance from the heme-iron, which may correspond to a glycerol molecule from the cryo-buffer, but the quality of the electron density did not allow unambiguous ligand identification. In addition, five PEG 400 molecules and three sulfate ions (SO_4^{2-}), originating from the crystallization solution, are present in the structure.

Structural comparison of CYP109A2 overall fold and active site architecture with related vitamin D_3 hydroxylases

CYP109A2 crystallized in an 'open' conformation, which is confirmed by structural alignments against other P450 structures deposited in the Protein Data Bank (PDB), giving the best match to those displaying open active site geometry (Table S2). The highest structural similarity is observed with the 'open' crystal structure of CYP109E1 containing four bound corticosterone molecules in the active site pocket (PDB entry 5L91, r.m.s.d. of 1.27 Å for 368 Cα atoms). In addition, a significant structural similarity is observed with the structures of Vdh from *P. autotrophica*, considering both its wild type and its mutant forms (L348M and F106V) [36] (e.g. wild type Vdh, PDB entry 3A4G, r.m.s.d. of 1.44 Å for 349 Cα atoms). Visual analysis of the CYP109A2, CYP109E1 and Vdh structures shows that they contain similar, wide-open, funnel-shaped and solvent exposed 'distal' heme pockets. The main difference concerns the BC loop/SRS1 region (loop connecting the B and C helices, residues 71-87 in CYP109A2), whose length, fold and sequence vary significantly in the

structures of the three different enzymes (**Figures 2 and 7B**). CYP109A2 contains the shortest BC loop, whereas the longest loop is present in Vdh (**Figure 2**). While the BC loop in CYP109A2 folds towards the active site pocket, the loops in CYP109E1 and especially Vdh are folded more outwards (**Figures 7B and 8A**). As a result, the BC loop in CYP109A2 restricts access of substrate to the heme-iron to a larger extent than the BC loops in CYP109E1 and Vdh. The structural differences in the BC loop region further result in a striking difference in the distribution of hydrophobic and polar regions in the active sites of the three P450s (**Figure 8B, C and D**). In the region closest to the heme, the active site pockets in CYP109A2, CYP109E1 and Vdh display a similar, mainly apolar surface due to the presence of conserved hydrophobic residues, i.e., Leu84, Met85 (SRS1); Leu237, Val240, Ala241 (SRS4); Ile288, Ile291 (SRS5) and Phe389, Val390 (SRS6) in CYP109A2. A bit further away from the heme, however, the surfaces show some significant differences. Residues Arg74, Glu78, Arg79 and Glu81 of the BC loop in CYP109A2 form a highly polar surface restricting one side of the active site pocket (**Figure 8B**). In the active site pocket of CYP109E1 the BC loop also forms a polar surface, but it is much less pronounced as in CYP109A2 (**Figure 8C**), while in Vdh it is nearly absent (**Figure 8D**). The observed differences in BC loop geometry, in the extent of restricted access to the heme-iron and in the apolar/polar surface distribution around the active site pocket will likely influence the substrate binding modes in the three different enzymes. In the absence of a closed, VD_3-bound crystal structure of CYP109A2 it remains unknown whether these differences are directly related to the differences in regio-selectivity for VD_3 hydroxylation. Arguably, a significantly more polar BC loop region and a more constricted active site pocket in CYP109A2, in comparison to CYP109E1 or Vdh, allows for lower occurrence of differently stabilized VD_3 binding modes, therefore resulting in an increased regio-selectivity of substrate hydroxylation in comparison to CYP109E1 or Vdh, as observed experimentally. Indeed, the BC loop, a highly variable and flexible region in P450s, has been previously associated with differences in substrate specificity [37].

Prediction of vitamin D_3 binding residues in CYP109A2

Unfortunately, our experiments to crystalize a structure of CYP109A2 in its closed conformation were not successful, despite extensive efforts to obtain a structure of substrate-bound and/or product-bound CYP109A2. The residues known or predicted to bind and stabilize VD_3 in the active sites of CYP109E1 or Vdh were therefore compared to their structural equivalents in CYP109A2 (Table 2). Based on this comparison the following thirteen residues of CYP109A2 are predicted to interact with VD_3: Arg74, Leu84, Met85 from SRS1; Leu167, Val168 from SRS2; Arg186, Asn187, Val190 from SRS3; Leu237, Val240, Thr245 from SRS4; Ile291 from SRS5 and Phe389 from SRS6 (**Figure 9**). Of these residues, six are identical to those in CYP109E1 or Vdh, while the other residues are relatively similar. This significant degree of conservation further supports the notion that the overall binding mode of VD_3 leading to 25-hydroxylation is similar in these enzymes. Whether the predicted residues in CYP109A2 indeed have important roles in substrate binding and conversion has to await further analysis by site-directed mutagenesis.

In summary, we have shown that cytochrome P450 monooxygenase CYP109A2 from *Bacillus megaterium* is able to catalyze hydroxylation of vitamin D_3 to its 25-hydroxylated derivative, with high regio-selectivity. The activity was reconstituted in the whole-cell based system, allowing to obtain high yields of the desired 25-hydroxylated product and therefore renders the system as most efficient and regio-selective in comparison to other related bacterial P450s previously shown to oxidize vitamin D_3. These characteristics highlight the significance of CYP109A2 as a promising, potential industrially applicable biocatalyst. The crystal structure of CYP109A2 allowed for a detailed analysis of the active site geometry and prediction of VD_3 interacting residues. Our study forms the basis to guide future efforts to explore the substrate scope and selectivity of CYP109A2.

Experimental Procedures

Materials

VD_3, 25(OH)VD_3, (2-hydroxypropyl)-β-cyclodextrin and saponin (from quillaja bark) were purchased from Sigma-Aldrich Chemie GmbH (Steinheim, Germany). Isopropyl β-D-1-thiogalactopyranoside (IPTG) and δ-aminolevulinic acid (δ-ALA) were purchased from Carbolution Chemicals GmbH (Saarbruecken, Germany). Bacterial media were purchased from Becton Dickinson (Heidelberg, Germany). All other chemicals were from standard sources and of highest purity available. The *E. coli* TOP10 strain used for cloning purposes was bought from Invitrogen, whereas the *E. coli* C43 (DE3) strain and pET17b (Merck Bioscience, Bad Soden, Germnay) vector used for heterologous expression were purchased from Novagen. The pSMF2.1 vector was used for protein expression in *B. megaterium* MS94 [38].

Cloning

The *cyp109a2* gene (GenBank GeneID 9117426) was amplified by polymerase chain reaction (PCR) using genomic DNA of *B. megaterium* MS941 as template (mutant derived from DSM319 strain) [39,40]. To allow expression of CYP109A2 carrying a C-terminal His_6-tag, the *cyp109a2* gene was extended with a sequence at the 3`end coding for the desired tag and cloned into the expression vector pET17b with the *Nde*I/*Kpn*I restriction sites yielding pET17b.CYP109A2. The primer pair used was 5´-ATATCATATGAATCCAAAAGCAGT-GAAAAGAGAAAATCG-3´ and 5´-TTAGGTACCTCAATGGTGGTGGTGATGATG-GGCGTTATTAAAAGCAATTTC-3´. The *cyp109a2* gene was also cloned into the pSMF2.1 vector using *Pac*I/*Spe*I restriction sites and including an optimized ribosomal binding site for *B. megaterium*, yielding pSMF2.1.CYP109A2. The primer pairs used were 5´-CGCTTAATTAAAAATCAAGGAGGTGAATGTACAATGAATCCAAAAGCAGTGAAAAGAG-3´ and 5`-TTATCAACTAGTTCAGGCGTTATTAAAAGCAATTTCATA-TTTTTTTACC-3´.

Heterologous gene expression and purification

The *E. coli* C43 (DE3) cells serving as expression host were transformed with the pET17b.CYP109A2 construct and cultured overnight in Luria-Bertani (LB) medium containing 100 μg mL^{-1} ampicillin, at 37 °C and 140 rpm. The overnight culture was used to inoculate a 250 mL main culture (Terrific Broth (TB) medium) containing 100 μg mL^{-1} ampicillin. The main culture was grown at 37 °C and 140 rpm until the optical density at 600 nm reached 0.6. Gene expression was induced by addition of 1 mM IPTG and 0.5 mM δ-ALA (heme precursor). The culture was further grown at 30 °C (120 rpm) for 24 h. The *E. coli* cells were harvested by centrifugation at 4,500 rpm for 30 min, and cell pellets were stored at − 20 °C until purification. For biochemical characterization, the protein was purified using a simplified procedure (one-step immobilized metal ion affinity chromatography). For crystallization, a three-step purification protocol was followed, as described previously (ion exchange chromatography, gel filtration, mixed-mode ion exchange chromatography) [18]. Purification of the redox partner protein pairs (BmCPR-Fdx2 and AdR-$Adx_{4\text{-}108}$) was performed according to the previously reported procedures [19,23].

Protein quality analysis

Spectroscopic properties of the purified P450 were analyzed using a double-beam spectrophotometer (UV-2101PC, Shimadzu, Japan). All UV-visible absorbance spectra were recorded from 200 to 700 nm. Concentration of CYP109A2 was estimated based on CO-difference spectroscopy (using an extinction coefficient of 91 $mm^{-1}\cdot cm^{-1}$), according to the method of Omura and Sato [41]. Protein purity and heterogeneity were analyzed by SDS-PAGE and dynamic light scattering (DLS), respectively. The dynamic light scattering measurement was performed using a DynaPro NanoStar instrument thermostated at 293 K (Wyatt Instruments, Santa Barbara, USA).

In vitro conversion and kinetic analysis

Enzymatic activity of CYP109A2 was reconstituted in an *in vitro* system containing 1 µM CYP109A2, 3 µM ferredoxin reductase (BmCPR or AdR), 20 µM ferredoxin (Fdx2 or Adx_{8-104}), 1 mM $MgCl_2$, and a cofactor regenerating system with glucose-6-phosphate (5 mM) and glucose-6-phosphate dehydrogenase (1 U), all in 50 mM potassium phosphate buffer containing 20% glycerol (pH 7.4) and at 250 µL final volume. VD_3 was dissolved in (2-hydroxypropyl)-β-cyclodextrin (2.25%) and added at 200 µM final concentration. The reaction was started by addition of 0.5 mM NADPH at 30 °C. After 1 h, the reaction was stopped and extracted twice with 500 µL of ethyl acetate. The organic phases were combined, evaporated to dryness and prepared for analysis by high-performance liquid chromatography (HPLC). For determination of kinetic parameters of VD_3 conversion by CYP109A2, three independent *in vitro* reactions were performed (each continuing for 20 min). VD_3 concentrations ranged from 0 to 400 µM. The k_{cat} and K_M values were calculated by plotting the product formation rate against increasing substrate concentration. The data were fit to the Michaelis-Menten equation using Origin software from OriginLab Corporation.

High-performance liquid chromatography (HPLC)

The HPLC analysis was carried out on a Jasco system (Pu-980 HPLC pump, AS-950 sampler, UV-975 UV/visible detector, LG-980-02 gradient unit; Jasco, Gross-Umstadt, Germany) equipped with a reversed-phase Nucleodur 100-5 C_{18} column (125 x 4 mm; Macherey-Nagel, Düren, Germany). The column temperature was adjusted to 40 °C. The samples were dissolved in 200 µL of acetonitrile. A linear gradient of 60–100% aqueous acetonitrile was applied for 15 min (1 mL· min^{-1} flow rate), followed by 100% acetonitrile for 15 min. The UV detection of the substrate and products was accomplished at 265 nm. Assuming that the absorption properties of the products did not differ from the substrate, the product formation was calculated as the relative peak area (area %) of the HPLC chromatograms, dividing each respective product peak area by the sum of products and substrate areas.

Whole-cell conversion in *B. megaterium* MS941

The whole-cell conversions were performed in *B. megaterium* MS941 cells transformed with the pSMF2.1.CYP109A2 construct using the polyethylene glycol-mediated protoplast transformation method [42]. Initial seed culture was prepared in LB medium (10 µg·mL^{-1} tetracycline). The main culture (50 mL TB medium, 10 µg·mL^{-1} tetracycline) was inoculated with 500 µL of the seed culture (dilution 1:100) in 300 mL baffled flasks and incubated under shaking at 37 °C (140 rpm). The culture was grown to OD_{578} of 0.4 and at that point the recombinant gene expression was induced with xylose (5 mg·mL^{-1}). The culture was further grown at 30°C for 24h (140 rpm). To initialize the whole-cell conversion experiments, after 24 h of protein expression the substrate (200 µM VD_3 dissolved in 45% (2-hydroxypropyl)-β-cyclodextrin)) and a membrane-solubilizing agent (4% *Quillaja* Saponin) were added to the cultures. Afterwards the conversions were allowed to proceed at 30 °C and 120 rpm for 24 h. Finally, 500 µL samples of the cultures were taken, extracted and prepared for HPLC analysis.

Product isolation and NMR characterization

In order to obtain sufficient amounts of the reaction product, the whole-cell based cultivation and conversion experiments were scaled up to 400 mL (2 x 200 mL) and performed as described above. The main reaction product (P1) was isolated and purified by reversed-phase HPLC using a preparative column VP Nucleodur 100-5 C_{18} (8.0 x 250 mm, Macherey-Nagel, Düren, Germany). The NMR spectra were recorded in $CDCl_3$ with a Bruker Avance 500 NMR spectrometer at 298 K. The chemical shifts were relative to $CHCl_3$ at δ 7.26 (^{1}H NMR) and $CDCl_3$ at δ 77.00 (^{13}C NMR) using the standard δ notation in parts per million. The 1D NMR (^{1}H and ^{13}C NMR) and the 2D NMR spectra (gs-HH-COSY and gs-HSQCED) were recorded using the BRUKER pulse program library.

Crystallization, data collection and structure determination

Aliquots with purified CYP109A2 were thawed, buffer exchanged to 20 mM Tris/HCl, pH 8.0, 0.1 mM dithioerythritol and concentrated to 40 $mg \cdot mL^{-1}$ by ultrafiltration using a 30-kDa cutoff membrane. Crystallization condition screening was done in 96-well plate format. Sitting-drop vapor-diffusion experiments were set up by a Mosquito crystallization robot (TTP LabTech), using commercially available high-throughput crystallization screens. Multiple hits were obtained overnight and the most promising condition (2.0 M ammonium sulfate, 0.1 M HEPES, pH 7.5, and 2% (v/v) PEG 400) was chosen for further optimization. Crystal optimization was performed manually at 293 K by the sitting-drop vapour diffusion method in 24-well plates using reservoir solution volumes of 300 µL. Protein drops were prepared by mixing 1 µL of protein and 1 µL of reservoir solution. A single, red plate-like crystal of CYP109A2 appeared after 3 days using a reservoir solution containing 1.9 M ammonium sulfate, 0.1 M HEPES, pH 7.5, and 2% (*v/v*) PEG 400 and needed about a week to grow to its final dimensions (0.12 x 0.04 x 0.03 mm^3).

For data collection, the crystal was shortly swiped through a cryobuffer droplet (mother liquor supplemented with 30% (*v/v*) glycerol) and flash-cooled to 100 K in a cold nitrogen gas cryostream. X-ray diffraction data were collected at the ID23-2 beam line of the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. Reflections were indexed and integrated using XDS [43], while scaling and merging of the data was done with AIMLESS from the CCP4 software suite [44]. The structure of CYP109A2 was determined by molecular replacement using Phaser from the PHENIX software package [45], applying the structure of substrate-free CYP109E1 (PDB entry 5L90**,** [18]) as a search model. The structure was completed with several cycles of model refinement using phenix.refine [46] alternated with manual model re-building using COOT [47]. At the final stages of refinement and model building water molecules were added to the model based on positive difference electron density peaks and strict interaction criteria. The final model was validated with MolProbity [48]. The model coordinates and structure factors are deposited in the Protein Data Bank with accession code 5OFQ.

Sequence and structure analysis

Protein sequences were compared using the Basic Local Alignment Search Tool (BLAST, NCBI). Substrate recognition sites (SRS) in CYP109A2 were identified by alignment with $P450_{cam}$ (WP_032492633) and followed a description defined by Gotoh [49]: SRS1: 71-87 (BC loop), SRS2: 163-169 (C' terminal part of F helix), SRS3: 185-192 (N' terminal part of G helix), SRS4: 229-247 (central part of I helix), SRS5: 284-293 (K-β5 connection) and SRS6: 384-392 (β9-β10 turn). Pairwise structural alignments were done using PDBeFold [50]. Multiple sequence alignment was done with Clustal Omega [51] and the output figure was prepared with ESPript3 [52]. The GenBank accession codes for analyzed sequences are as follows: CYP109A2 from *B. megaterium* DSM319 (WP_013082916), CYP109E1 from *B. megaterium* DSM319 (WP_013084555), CYP105A1 from *Streptomyces griseolus* (BAG50411), CYP105A2 from *Pseudonocardia autotrophica* (BAA05541), Vdh (CYP107BR1) from *P. autotrophica* **(**C4B644), CYP107CB2 from *B. lehensis* G1 (AIC83164).

Acknowledgements

This work was financially supported by the German Federation of Industrial Research Associations (AIF/ZIM project FKZ 2214512AJA) and the People Programme (Marie Curie Actions) of the European Union's 7th Framework Programme (FP7/2007-2013) under REA Grant Agreement 289217 (P4FIFTY). Thanks to Dr. Josef Zapp for measuring the NMR samples. Thanks to Dr. Flora Kiss and Dr. Mohammed Milhim for protein purification. Authors would also like to thank the European Synchrotron Radiation Facility (ESRF) and their staff for beam line access and data collection assistance.

Author Contributions

AA, IKJ, FH, AMWHT and RB designed the study. AA performed all biochemical experiments and collected data. EB purified the products for NMR analysis. AA, RB and FH analyzed and interpreted

the biochemical data. IKJ crystallized the protein, collected crystallographic data and determined the crystal structure. IKJ and AMWHT analyzed and interpreted the structural data. AA and IKJ wrote the manuscript. AMWHT, RB and FH provided supervision and participated in writing of the manuscript. All authors read and approved the final version of the manuscript.

Supporting Information

Table S1: NMR data for the main product of CYP109A2-catalyzed VD_3 conversion.

Table S2: Highest ranked structural homologues of CYP109A2 found in the PDB.

References

1. Pludowski P, Holick MF, Grant WB, Konstantynowicz J, Mascarenhas MR, Haq A, Povoroznyuk V, Balatska N, Barbosa AP, Karonova T, Rudenka E, Misiorowski W, Zakharova I, Rudenka A, Łukaszkiewicz J, Marcinowska-Suchowierska E, Łaszcz N, Abramowicz P, Bhattoa HP & Wimalawansa SJ (2017) Vitamin D supplementation guidelines. *J Steroid Biochem Mol Biol*, doi: 10.1016/j.jsbmb.2017.01.021

2. Wacker M & Holick MF (2013) Sunlight and Vitamin D: A global perspective for health. *Dermato-endocrinology* **5**, 51-108.

3. Hossein-nezhad A & Holick MF (2013) Vitamin D for health: a global perspective. *Mayo Clin Proc* **88**, 720-755.

4. Jean G, Terrat JC, Vanel T, Hurot JM, Lorriaux C, Mayor B & Chazot C (2008) Daily oral 25-hydroxycholecalciferol supplementation for vitamin D deficiency in haemodialysis patients: effects on mineral metabolism and bone markers. *Nephrol Dial Transplant* **23**, 3670-3676.

5. Leichtmann GA, Bengoa JM, Bolt MJ & Sitrin MD (1991) Intestinal absorption of cholecalciferol and 25-hydroxycholecalciferol in patients with both Crohn's disease and intestinal resection. *Am J Clin Nutr* **54**, 548-552.

6. Zhu G & Okamura WH (1995) Synthesis of Vitamin D (Calciferol). *Chem Rev* **95**, 1877-1952.

7. Jones G, Prosser DE & Kaufmann M (2014) Cytochrome P450-mediated metabolism of vitamin D. *J Lipid Res* **55**, 13-31.

8. Bernhardt R (2006) Cytochromes P450 as versatile biocatalysts. *J Biotechnol* **124**, 128-145.

9. Bernhardt R & Urlacher VB (2014) Cytochromes P450 as promising catalysts for biotechnological application: chances and limitations. *Appl Microbiol Biotechnol* **98**, 6185-6203.

10. Girvan HM & Munro AW (2016) Applications of microbial cytochrome P450 enzymes in biotechnology and synthetic biology. *Curr Opin Chem Biol* **31**, 136-145.

11. Sasaki J, Mikami A, Mizoue K & Omura S (1991) Transformation of 25- and 1 alpha-hydroxyvitamin D3 to 1 alpha, 25-dihydroxyvitamin D3 by using *Streptomyces* sp. strains. *Appl Environ Microbiol* **57**, 2841-2846.

12. Fujii Y, Kabumoto H, Nishimura K, Fujii T, Yanai S, Takeda K, Tamura N, Arisawa A & Tamura T (2009) Purification, characterization, and directed evolution study of a vitamin D3 hydroxylase from *Pseudonocardia autotrophica*. *Biochem Biophys Res Commun* **385**, 170-175.

13. Abdulmughni A, Jóźwik IK, Putkaradze N, Brill E, Zapp J, Thunnissen AW, Hannemann F & Bernhardt R (2016) Characterization of cytochrome P450 CYP109E1 from *Bacillus megaterium* as a novel vitamin D_3 hydroxylase. *J Biotechnol* **243**, 38-47.

14. Hayashi K, Yasuda K, Sugimoto H, Ikushiro S, Kamakura M, Kittaka A, Horst RL, Chen TC, Ohta M, Shiro Y & Sakaki T (2010) Three-step hydroxylation of vitamin D3 by a genetically engineered CYP105A1: enzymes and catalysis. *FEBS J* **277**, 3999-4009.

15. Hayashi K, Sugimoto H, Shinkyo R, Yamada M, Ikeda S, Ikushiro S, Kamakura M, Shiro Y & Sakaki T (2008) Structure-based design of a highly active vitamin D hydroxylase from *Streptomyces griseolus* CYP105A1. *Biochemistry* **47**, 11964-11972.

16. Yasutake Y, Fujii Y, Nishioka T, Cheon WK, Arisawa A & Tamura T (2010) Structural evidence for enhancement of sequential vitamin D3 hydroxylation activities by directed evolution of cytochrome P450 vitamin D3 hydroxylase. *J Biol Chem* **285**, 31193-31201.

17. Yasutake Y, Nishioka T, Imoto N & Tamura T (2013) A single mutation at the ferredoxin binding site of P450 Vdh enables efficient biocatalytic production of 25-hydroxyvitamin D(3). *Chembiochem* **14**, 2284-2291.

18. Jóźwik IK, Kiss FM, Gricman Ł, Abdulmughni A, Brill E, Zapp J, Pleiss J, Bernhardt R & Thunnissen AW (2016) Structural basis of steroid binding and oxidation by the cytochrome P450 CYP109E1 from *Bacillus megaterium*. *FEBS J* **283**, 4128-4148.

19. Brill E, Hannemann F, Zapp J, Bruning G, Jauch J & Bernhardt R (2014) A new cytochrome P450 system from *Bacillus megaterium* DSM319 for the hydroxylation of 11-keto-beta-boswellic acid (KBA). *Appl Microbiol Biotechnol* **98**, 1701-1717.

20. Hamdane D, Zhang H & Hollenberg P (2008) Oxygen activation by cytochrome P450 monooxygenase. *Photosynth Res* **98**, 657-666.

21. Kiss FM, Schmitz D, Zapp J, Dier TK, Volmer DA & Bernhardt R (2015) Comparison of CYP106A1 and CYP106A2 from *Bacillus megaterium* - identification of a novel 11-oxidase activity. *Appl Microbiol Biotechnol* **99**, 8495-8514.

22. Hannemann F, Bichet A, Ewen KM & Bernhardt R (2007) Cytochrome P450 systems—biological variations of electron transport chains. *Biochim Biophys Acta* **1770**, 330-344.

23. Milhim M, Gerber A, Neunzig J, Hannemann F & Bernhardt R (2016) A novel NADPH-dependent flavoprotein reductase from *Bacillus megaterium* acts as an efficient cytochrome P450 reductase. *J Biotechnol* **231**, 83-94.

24. Sakaki T, Sugimoto H, Hayashi K, Yasuda K, Munetsuna E, Kamakura M, Ikushiro S & Shiro Y (2011) Bioconversion of vitamin D to its active form by bacterial or mammalian cytochrome P450. *Biochim Biophys Acta* **1814**, 249-256.

25. Sawada N, Sakaki T, Yoneda S, Kusudo T, Shinkyo R, Ohta M & Inouye K (2004) Conversion of vitamin D3 to 1alpha,25-dihydroxyvitamin D3 by *Streptomyces griseolus* cytochrome P450SU-1. *Biochem Biophys Res Commun* **320**, 156-164.

26. Bunk B, Schulz A, Stammen S, Munch R, Warren MJ, Rohde M, Jahn D & Biedendieck R (2010) A short story about a big magic bug. *Bioeng Bugs* **1**, 85-91.

27. Vary PS, Biedendieck R, Fuerch T, Meinhardt F, Rohde M, Deckwer WD & Jahn D (2007) *Bacillus megaterium*--from simple soil bacterium to industrial protein production host. *Appl Microbiol Biotechnol* **76**, 957-967.

28. Gerber A, Kleser M, Biedendieck R, Bernhardt R & Hannemann F (2015) Functionalized PHB granules provide the basis for the efficient side-chain cleavage of cholesterol and analogs in recombinant *Bacillus megaterium*. *Microb Cell Fact* **14**, 107-015-0300-y.

29. Putkaradze N, Kiss FM, Schmitz D, Zapp J, Hutter MC & Bernhardt R (2017) Biotransformation of prednisone and dexamethasone by cytochrome P450 based systems - Identification of new potential drug candidates. *J Biotechnol* **242**, 101-110.

30. Ehrhardt M, Gerber A, Hannemann F & Bernhardt R (2016) Expression of human CYP27A1 in *B. megaterium* for the efficient hydroxylation of cholesterol, vitamin D3 and 7-dehydrocholesterol. *J Biotechnol* **218**, 34-40.

31. Zhang W, Liu Y, Yan J, Cao S, Bai F, Yang Y, Huang S, Yao L, Anzai Y, Kato F, Podust LM, Sherman DH & Li S (2014) New reactions and products resulting from alternative interactions between the P450 enzyme and redox partners. *J Am Chem Soc* **136**, 3640-3646.

32. Khatri Y, Schifrin A & Bernhardt R (2017) Investigating the effect of available redox protein ratios for the conversion of a steroid by a myxobacterial CYP260A1. *FEBS Lett* **591**, 1126-1140.

33. Helmer B, Schnoes HK & DeLuca HF (1985) 1H nuclear magnetic resonance studies of the conformations of vitamin D compounds in various solvents. *Arch Biochem Biophys* **241**, 608-615.

34. Mizhiritskii MD, Konstantinovskii LE & Vishkautsan R (1996) 2D NMR study of solution conformations and complete 1H and 13C chemical shifts assignments of vitamin D metabolites and analogs. *Tetrahedron* **52**, 1239-1252.

35. Takeda K, Asou T, Matsuda A, Kimura K, Okamura K, Okamoto R, Sasaki J, Adachi T & Omura S (1994) Application of cyclodextrin to microbial transformation of vitamin D3 to 25-hydroxyvitamin D3 and 1α,25-dihydroxyvitamin D3. *J Ferment Bioeng* **78**, 380-382.

36. Yasutake Y, Kameda T & Tamura T (2017) Structural insights into the mechanism of the drastic changes in enzymatic activity of the cytochrome P450 vitamin D 3 hydroxylase (CYP107BR1) caused by a mutation distant from the active site. *Acta Crystallogr. Sect. F* **73**, 266–275.

37. Lepesheva GI, Virus C & Waterman MR (2003) Conservation in the CYP51 family. Role of the B' helix/BC loop and helices F and G in enzymatic function. *Biochemistry* **42**, 9091-9101.

38. Bleif S, Hannemann F, Zapp J, Hartmann D, Jauch J & Bernhardt R (2012) A new *Bacillus megaterium* whole-cell catalyst for the hydroxylation of the pentacyclic triterpene 11-keto-beta-boswellic acid (KBA) based on a recombinant cytochrome P450 system. *Appl Microbiol Biotechnol* **93**, 1135-1146.

39. Wittchen KD & Meinhardt F (1995) Inactivation of the major extracellular protease from *Bacillus megaterium* DSM319 by gene replacement. *Appl Microbiol Biotechnol* **42**, 871-877.

40. Eppinger M, Bunk B, Johns MA, Edirisinghe JN, Kutumbaka KK, Koenig SS, Creasy HH, Rosovitz MJ, Riley DR, Daugherty S, Martin M, Elbourne LD, Paulsen I, Biedendieck R, Braun C, Grayburn S, Dhingra S, Lukyanchuk V, Ball B, Ul-Qamar R, Seibel J, Bremer E, Jahn D, Ravel J & Vary PS (2011) Genome sequences of the biotechnologically important *Bacillus megaterium* strains QM B1551 and DSM319. *J Bacteriol* **193**, 4199-4213.

41. Omura T & Sato R (1964) The carbon monoxide-binding pigment of liver microsomes: I. Evidence for its hemoprotein nature. *J Biol Chem* **239**, 2370-2378.

42. Barg H, Malten M, Jahn M & Jahn D (2005) Protein and vitamin production in *Bacillus megaterium*. In *Microbial Processes and Products* (Barredo, J., ed.), pp. 205-23. Humana Press, New York, NY.

43. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132.

44. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A & Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242.

45. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC & Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221.

46. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH & Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352-367.

47. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132.

48. Chen VB, Arendall WB,3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS & Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21.

49. Gotoh O (1992) Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem* **267**, 83-90.

50. Krissinel E & Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**, 2256-2268.

51. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD & Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539.

52. Robert X & Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-4.

Table 1. Crystallographic data collection and refinement statistics of CYP109A2.

	CYP109A2
PDB code	**5OFQ**
Model statistics	
Monomers in the AU	4
Solvent content (%)	50
Ligands	5 x PEG 400, 3 x sulfate ion
Data collection	
Beamline (ESRF)	ID23-2
Wavelength (Å)	0.873
Resolution range (Å)	52-2.70 (2.78-2.70)[a]
Space group	$P2_12_12_1$
Unit-cell parameters	
a, b, c (Å)	77.12 155.53 158.15
α, β, γ, °	90 90 90
Observed reflections	287,529 (25,387)
Unique reflections	51,368 (4,458)
Multiplicity	5.6 (5.7)
Completeness (%)	97.3 (98.5)
$\langle I/\sigma(I) \rangle$	12.4 (1.7)
R_{merge} (%)	0.10 (0.94)
$R_{p.i.m.}$	0.07 (0.63)
CC (1/2) (%)	99.6 (52.3)
Refinement	
R_{work} (%)	22.4
R_{free} (%)	27.5
R.m.s.deviation, bond lengths (Å)	0.018
R.m.s.deviation, bond angles (°)	1.418

Average B-factors (Å)^2	
Overall	50.5
Protein	50.4
Heme	50.0
Ramachandran plot statistics	
Most favoured (%)	95.6
Allowed regions (%)	4.4
Disallowed regions (%)	0
Molprobity overall score	1.49

[a] Values in parentheses are for the highest resolution shell.

Table 2. Identification of residues with predicted VD_3 binding roles in CYP109A2, based on structural comparison to CYP109E1 and Vdh. (**A**) Residues of CYP109E1 previously predicted to interact with the VD_3 molecule [13] and their structurally equivalent residues in CYP109A2. (**B**) VD_3-binding residues in the crystal structure of VD_3-bound Vdh (PDB id: 3A50) [16] and their structurally equivalent residues in CYP109A2. Residues of CYP109A2 appearing in both panels are highlighted in *blue*. The residues of CYP109E1 studied previously by site-directed mutagenesis [13,18] are highlighted in *grey*. Conserved CYP109A2 residues are shown in bold.

A	**Predicted VD_3-binding residues CYP109E1**	**CYP109A2**
SRS1	Arg69	**Arg74**
	Pro71	-
	Thr78	-
	Leu80	-
	Ile85	Leu84
	Asn86	Met85
SRS2	Ile168	Leu167
	Val169	**Val168**
SRS3	Asn191	Val190
SRS4	Leu238	**Leu237**
	Ile241	Val240
	Thr246	**Thr245**

B	**VD_3-binding residues Vdh (3A50)**	**CYP109A2**
SRS1	Thr84	-
	Met86	-
	Ile88	Leu84
	Leu89	Met85
SRS2	Leu171	**Leu167**
SRS3	Lys180	Arg186
	Asn181	**Asn187**
	Met184	Val190
SRS4	Leu232	**Leu237**
	Ile235	Val240
	Thr240	**Thr245**
SRS5	Pro287	Ile291
SRS6	Leu387	Phe389

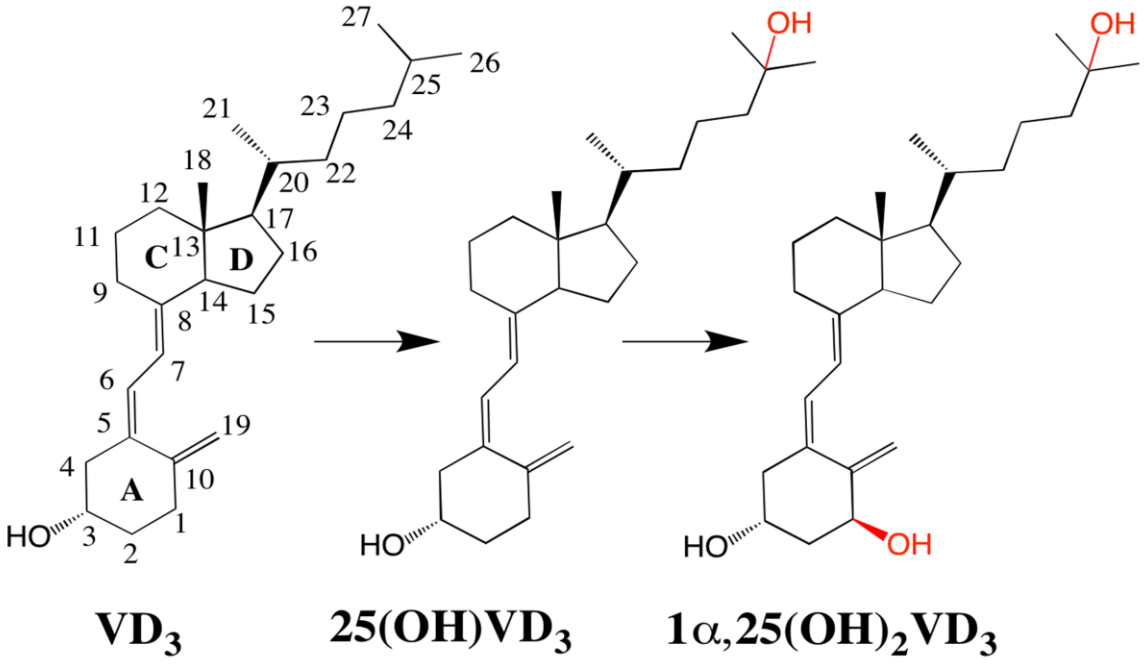


Figure 1: The activation pathway of vitamin D_3 in mammals. First, vitamin D_3 undergoes a C-25 hydroxylation resulting in the formation of $25(OH)VD_3$ (in the liver) followed by the second hydroxylation at C-1 atom producing $1\alpha,25(OH)_2VD_3$ (in the kidneys).

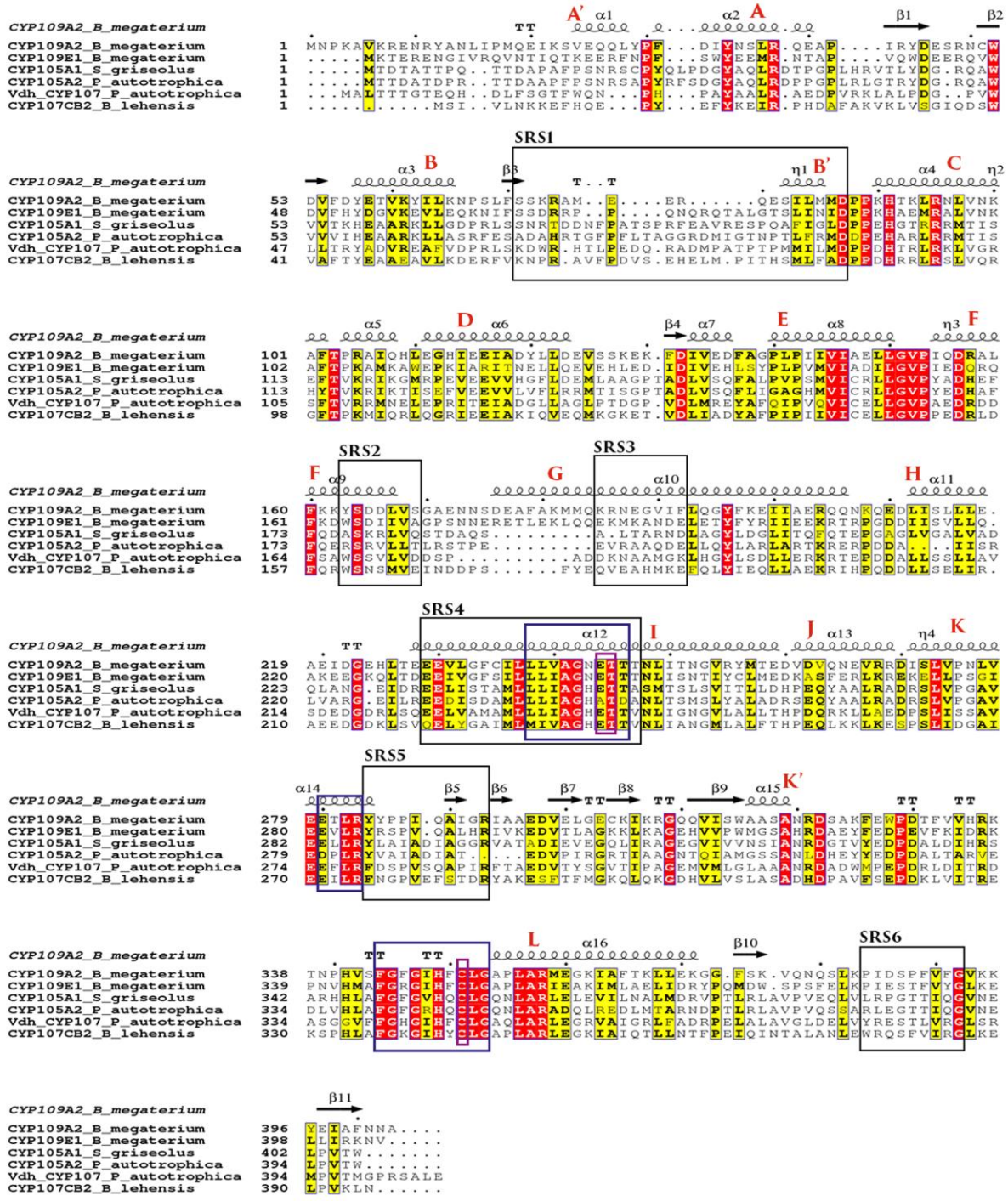


Figure 2. Multiple sequence alignment of CYP109A2 and other bacterial P450s showing 25-hydroxylation activity with VD_3 or VD_3 derivatives. Secondary structural elements are shown as in the crystal structure of CYP109A2. Conserved and similar residues are highlighted in red and yellow, respectively. Helices are labeled A-L. Residues in the six substrate recognition sites (SRS) are enclosed in black boxes. The characteristic P450 sequence motifs are in blue boxes: the heme-binding domain signature just before the L helix, central part of I helix with the highly conserved 'acid-alcohol pair' of residues, and the 'ExxR' motif in the K helix.

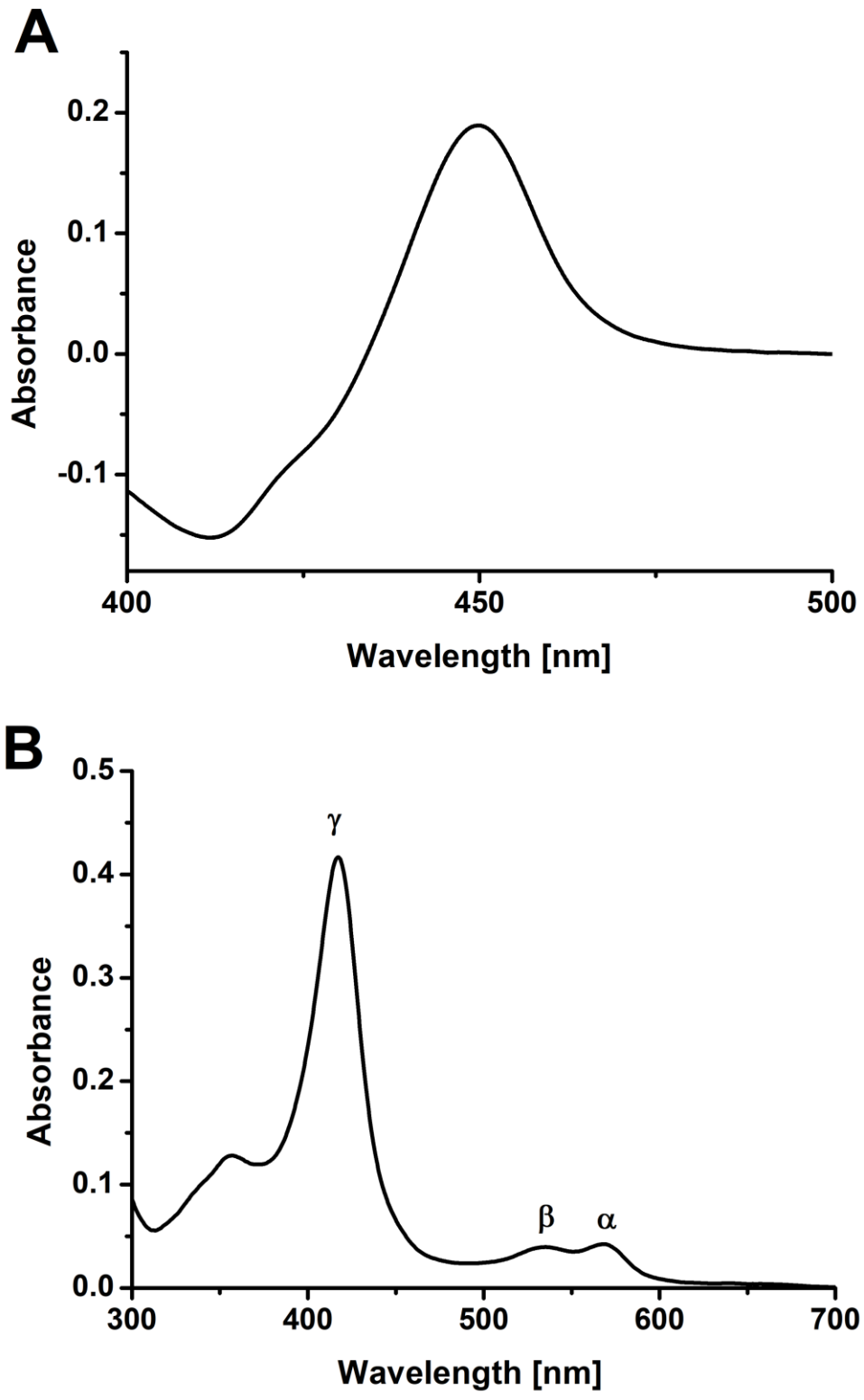


Figure 3. Spectroscopic properties of CYP109A2. (**A**) The UV-visible absorbance spectrum of the reduced CO-bound form of CYP109A2 showing the characteristic Soret peak at 450 nm. (**B**) The UV-visible **absorbance** spectrum of ligand-free CYP109A2 displaying typical peaks; γ (at 417 nm), α (568 nm) and β (536 nm).

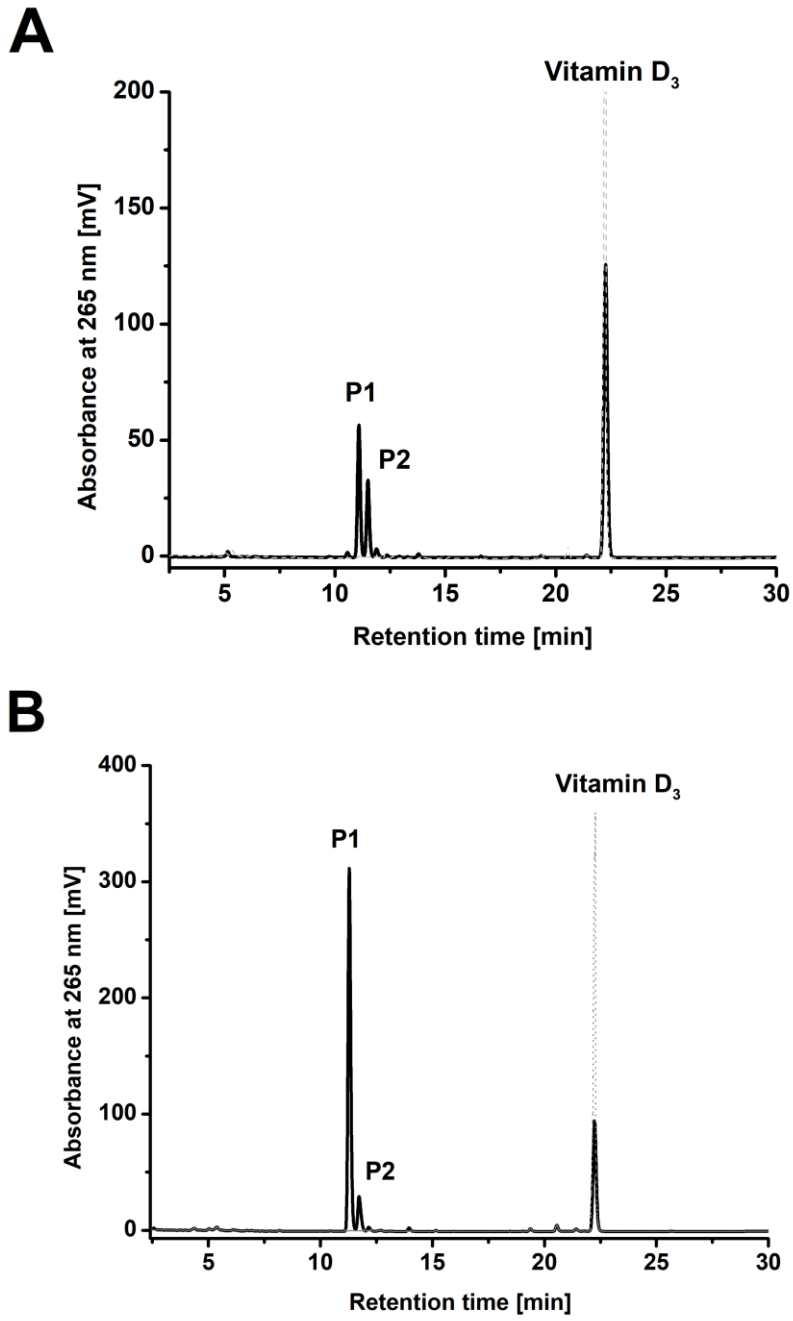


Figure 4. HPLC analysis of CYP109A2-dependent VD_3 conversion. (**A**) *In vitro* conversion. (**B**) Whole-cell conversion in *B. megaterium*. The products are labeled P1-P2. The authentic standard of VD_3 (grey) was detected with the same HPLC method.

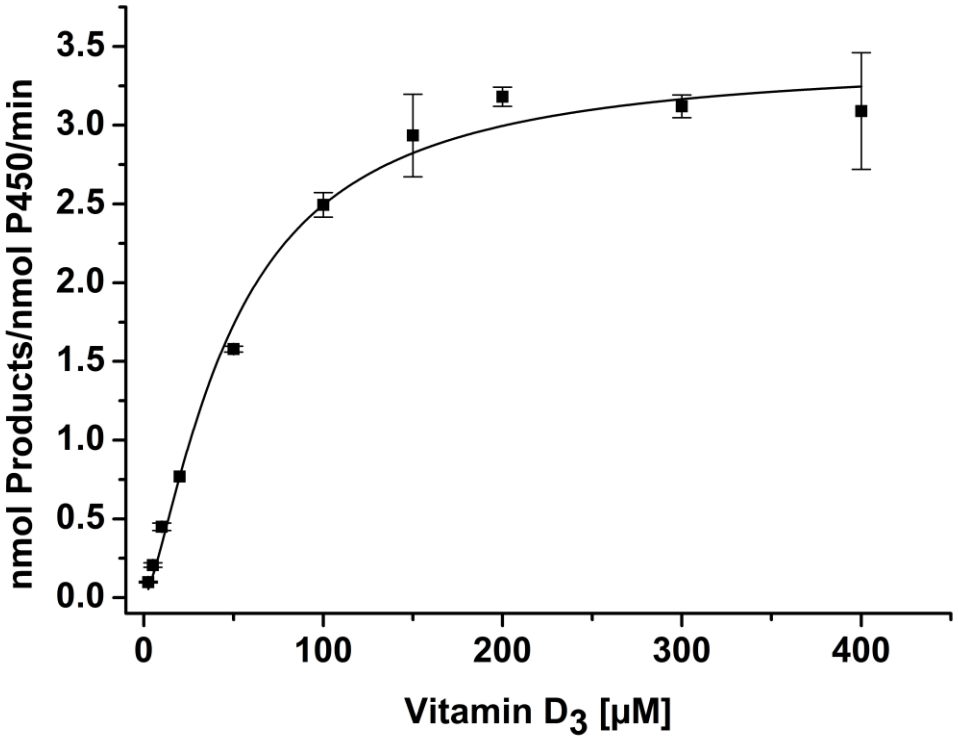


Figure 5. Determination of the kinetic parameters of VD_3 conversion catalyzed by CYP109A2. Mean values and standard deviations (error bars) were calculated from the results of three independent experiments. The data were fitted to the Michaelis-Menten equation.

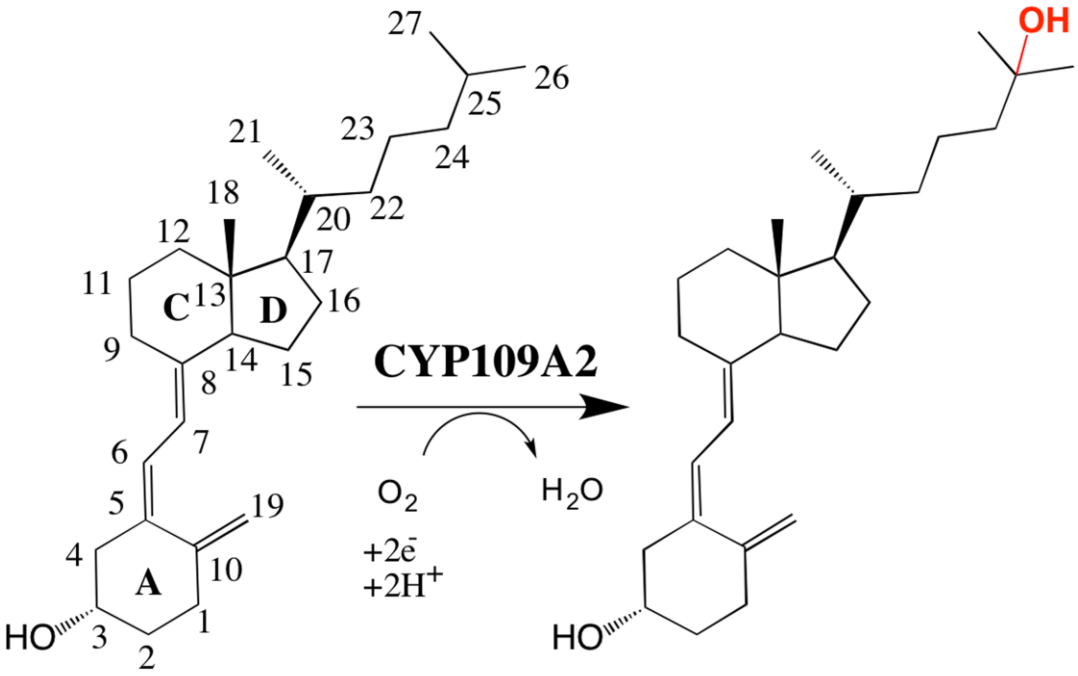
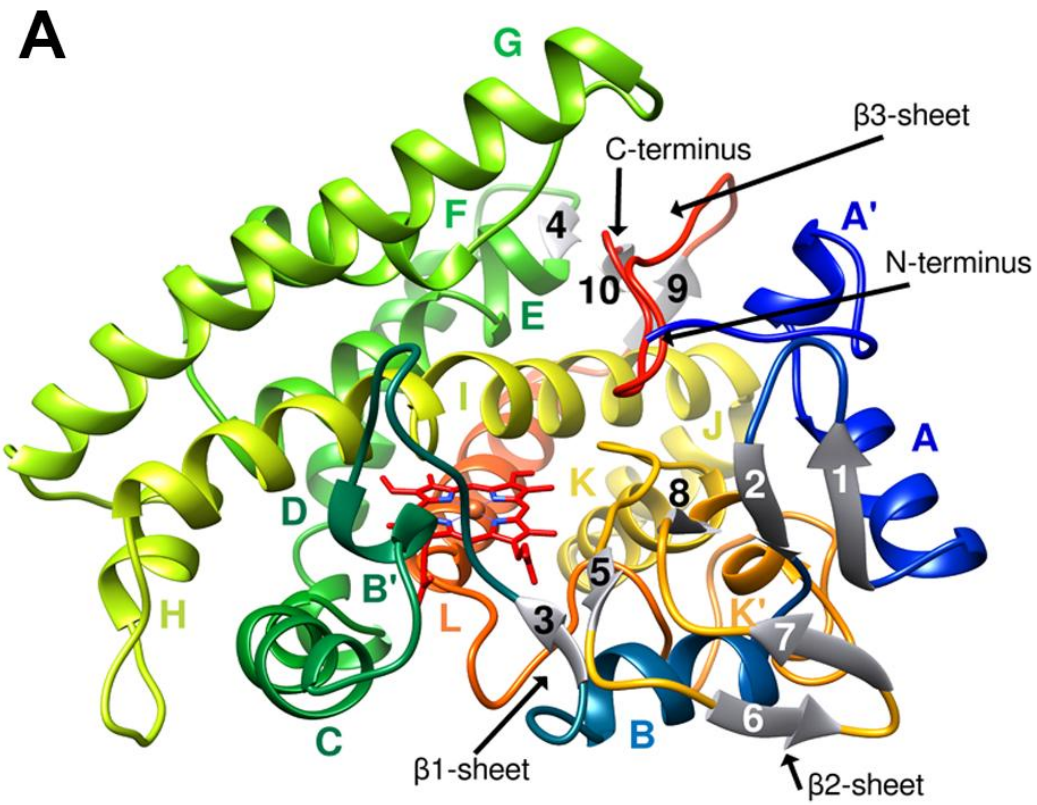


Figure 6. Scheme of the CYP109A2-catalysed conversion of VD_3. Hydroxylation site is indicated in red.



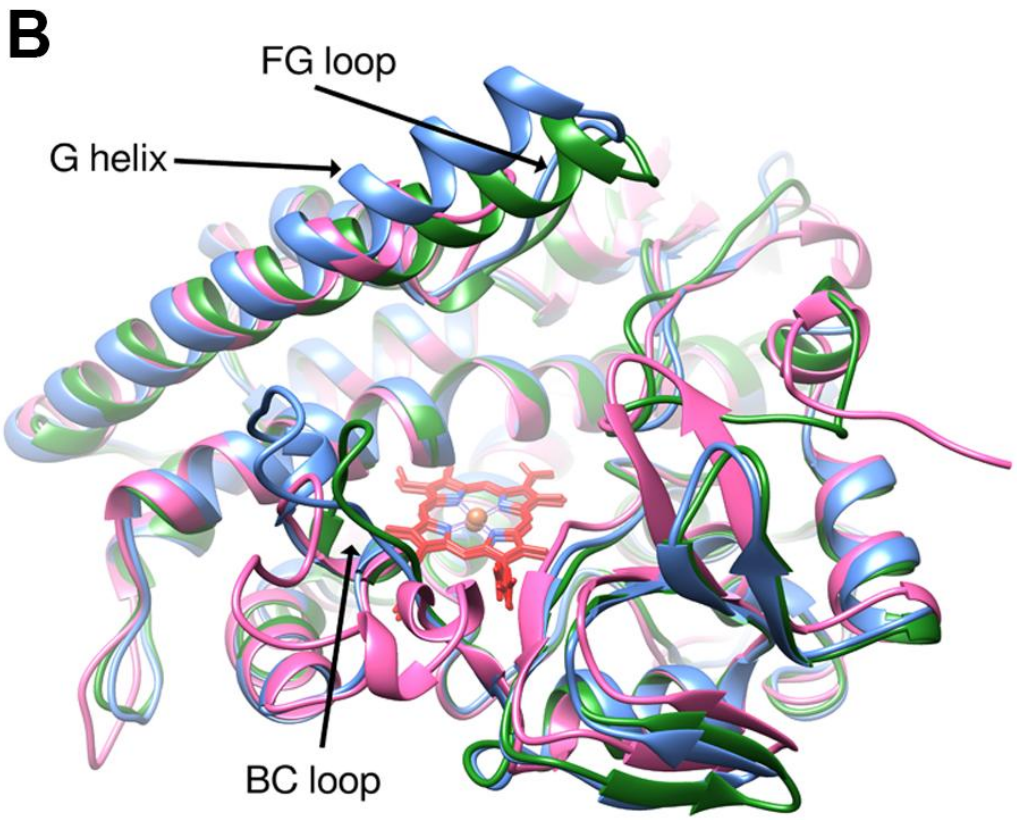


Figure 7. Crystal structure of substrate-free CYP109A2. (**A**) Ribbon representation with secondary structure elements labeled following the common P450 nomenclature (rainbow coloring from *blue* N' terminus to *red* C' terminus). Helices are labeled A-L following the common P450 nomenclature of secondary structure elements. Heme is shown as *red* stick model. (**B**) Superposition of the overall open conformations of CYP109A2, CYP109E1 from *B. megaterium* DSM319 and Vdh from *Pseudonocardia autotrophica*. Cartoon representation of the aligned structures of CYP109A2 (*green*), CYP109E1 (5L90, *blue*) and Vdh (3A4G, *pink*) is shown.

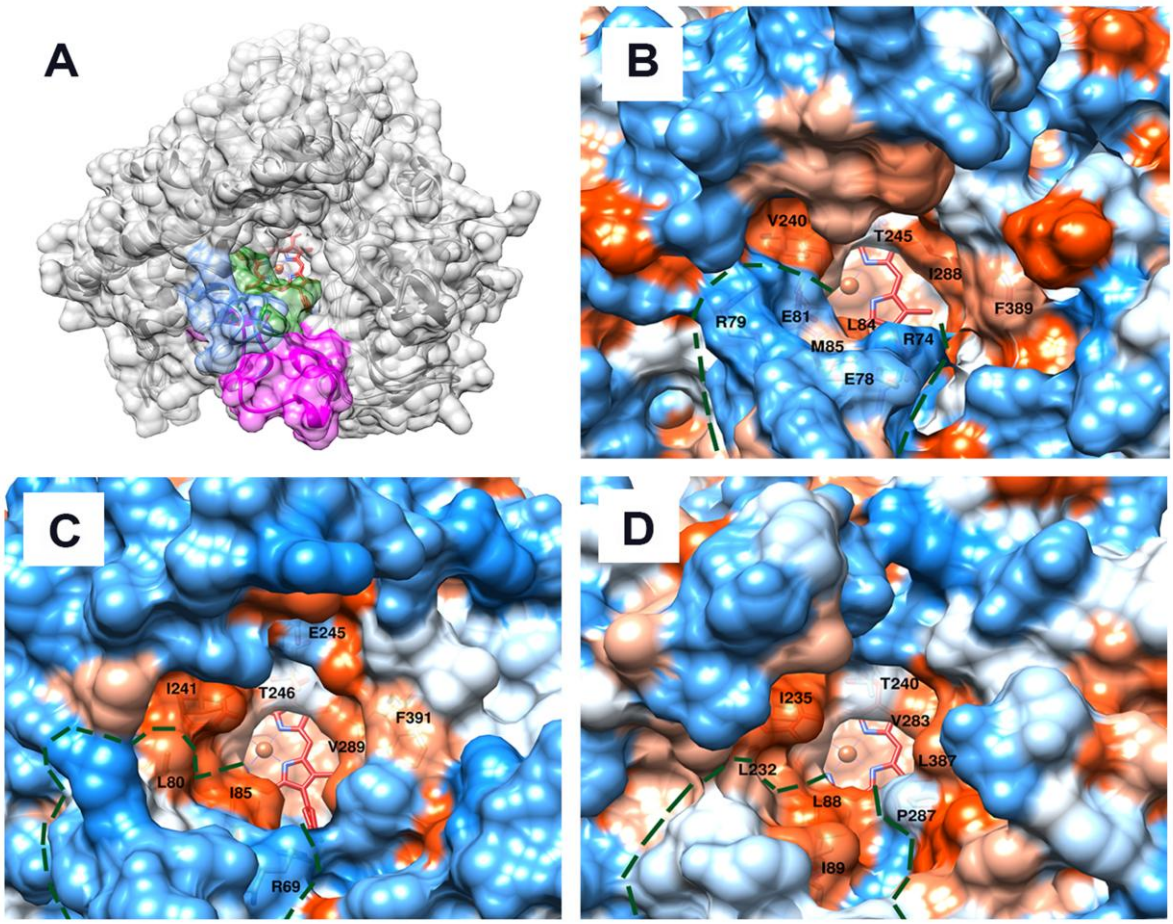


Figure 8. Comparison of the substrate binding pockets in CYP109A2, CYP109E1 from *B. megaterium* DSM319 and Vdh from *Pseudonocardia autotrophica*, all presenting an open conformation. (**A**) Three structures are superimposed and their molecular surfaces are shown in *grey*. The steric restriction of the active site pocket imposed by the BC loop region is significantly larger in the CYP109A2 structure (surface in *green*) than in CYP109E1 (5L90, in *blue*) and Vdh (3A4G, in *magenta*). (**B-D**) Surface representation of the 'distal' heme substrate binding pockets in CYP109A2 (**B**), CYP109E1 (**C**) and Vdh (**D**). Apolar residues are colored in *orange* and polar residues in *blue*. The different BC loop regions in the three P450s (lower parts of the images) are outlined with a *green* dashed line.

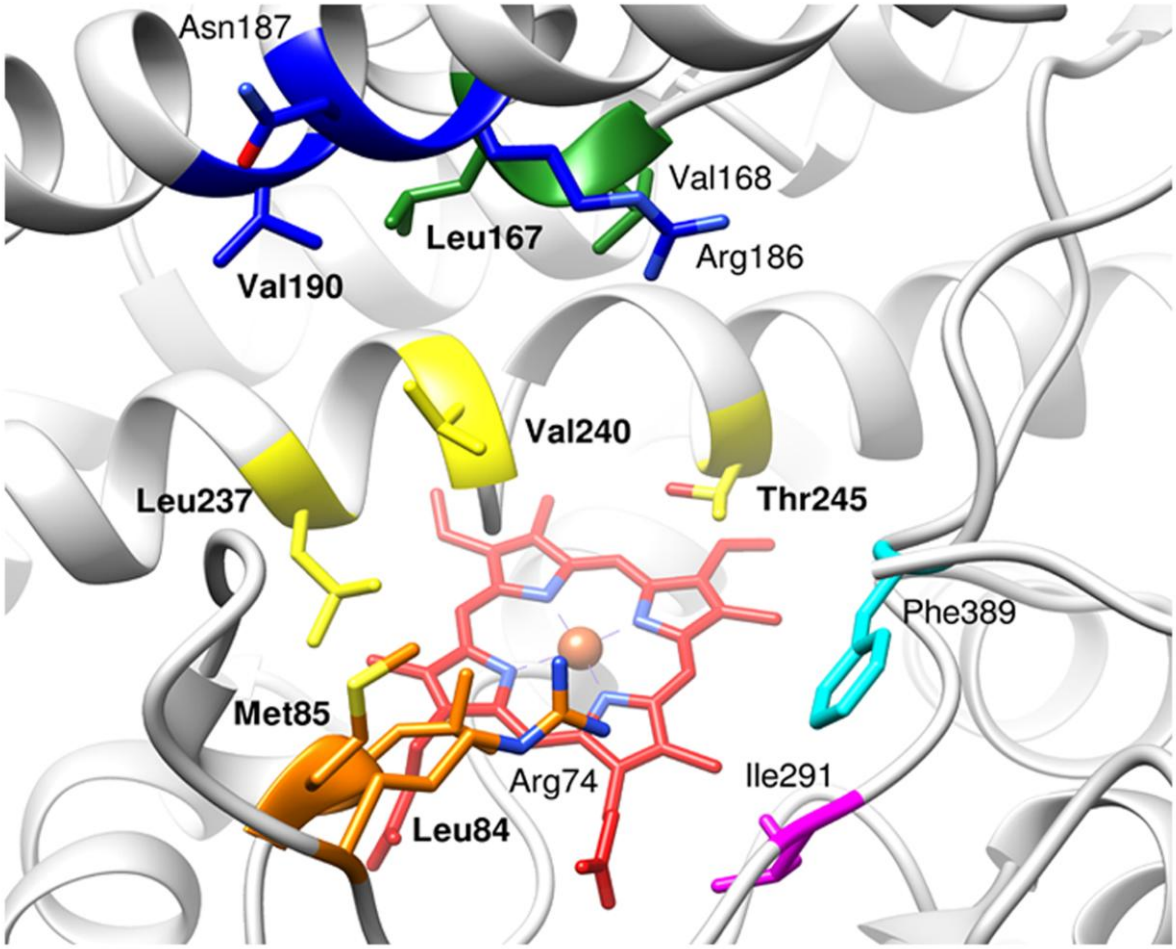


Figure 9. Predicted VD_3-binding residues in CYP109A2, identified based on structural comparison with CYP109E1 and Vdh (Table 2). The residues are shown as sticks and coloured according to the corresponding substrate recognition site (SRS1 *orange*, SRS2 *green*, SRS3 *blue*, SRS4 *yellow*, SRS5 *pink*, and SRS6 *cyan*). Residues overlapping in both predictions are labeled in bold. The heme group is shown as *red* coloured sticks.